

De huidige generatie digitale informatiedragers zal langzaam maar zeker uit de gratie raken, want niet duurzaam genoeg en met een beperkte opslagcapaciteit. Maar opslaan in DNA komt eraan. Hoewel de eerste toepassingen veelbelovend zijn, moeten er nog wel wat obstakels overwonnen worden.



Jasper Snoeren

Coördinator nieuwe media-collecties & papieren archieven bij het Nederlands Instituut voor Beeld en Geluid



De Data

Internetsnelheden nemen alsmat toe, en daarmee ook de hoeveelheid data die wereldwijd wordt gecreëerd. Iedere kattenvideo, iedere vakantiefoto en ieder boodschappenlijstje wordt digitaal bewaard en gedeeld. Daardoor zijn we haast ongemerkt voor een van de grootste uitdagingen in ons vakgebied komen te staan. Want waar de behoefte aan data-opslag blijft toenemen, zal de opslagcapaciteit van digitale dragers uiteindelijk tegen een maximum aanlopen. Deze dragers hebben bovendien een beperkte levensduur: magnetische tape gaat slechts dertig jaar mee, een harde schijf kun je al na tien jaar vervangen. Kortom, we hebben een nieuw hulpmiddel



sciencefiction voorbij opslag in DNA

nodig om data compacter en duurzamer op te slaan. Een oplossing blijken we al sinds het begin der tijden bij ons te dragen: DNA.

Veel voordelen

DNA (desoxyribonucleïnezuur) bevat alle erfelijke informatie van een organisme. Het vormt de blauwdruk voor de eiwitten die alle processen in het lichaam regelen. DNA heeft de handige eigenschap dat het enorme hoeveelheden complexe informatie kan opslaan in een extreem klein formaat. Als opslagmedium voor data heeft het dus een gigantisch potentieel: theoretisch kan er circa 215 petabyte per gram DNA worden bewaard. Dat is pakweg het

volledige archief van het Nederlands Instituut voor Beeld en Geluid in een rijstkorrel.

Maar niet alleen het formaat biedt voordelen, ook is het buitengewoon duurzaam. Het DNA uit fossielen van mammoeten of uit het lichaam van ijsmummie Ötzi is nog steeds leesbaar. Wanneer het koel en droog bewaard wordt, gaat DNA honderdduizenden jaren mee zonder dataverlies. Bovendien is het energie-efficiënt: de moleculen zelf verbruiken geen elektriciteit. Ook zal de kennis over het gebruik van DNA niet verdwijnen zolang de genetische informatie van de mens erin is opgeslagen. De techniek om het materiaal te maken, bewerken en ontcijferen zal

'De afgelopen vijftien jaar zijn als *proof of concept* al heel wat boeken, muziek en zelfs videoclipjes in DNA opgeslagen'

over honderd jaar ongetwijfeld verbeterd zijn, maar de informatie blijft onveranderd leesbaar.

Geen sciencefiction

Dat klinkt allemaal nog vrij futuristisch. Toch is DNA als opslagmedium niet langer sciencefiction. In de afgelopen vijftien jaar zijn grote stappen gezet: als *proof of concept* zijn inmiddels al heel wat boeken, muziek en zelfs videoclips in DNA opgeslagen. Zo hebben onderzoekers van de Universiteit van Washington en Microsoft honderd boeken van Project Gutenberg omgezet in DNA, evenals de Universele Verklaring van de Rechten van de Mens in alle talen. En niet te vergeten de geluids-

Muziek in een spuitbus

Om het twintigjarig bestaan van hun baanbrekende plaat *Mezzanine* te vieren, liet de Britse triphopgroep Massive Attack hun muziek vorig jaar opslaan in DNA. Het is daarmee het eerste album dat ooit op deze manier is vastgelegd. En daar bleef het niet bij: ze verkopen ook speciale graffiti waarin miljoenen DNA-kopieën van het album zijn opgelost in de zwarte verf.

[illegible]

Reald, ETH Zurich / Robert Grass

De 'genetische code' van het muziekalbum *Mezzanine*.

Dit deel vertegenwoordigt 1/27.000ste deel van de informatie op het gehele album

opnames op de gouden grammofoonplaat die in 1977 met de *Voyager* de ruimte in werd gezonden als boodschap aan buitenaardse levensvormen. Op dit moment zijn de onderzoekers bezig om de gehele Engelstalige Wikipedia-database vast te leggen.

Hoe werkt het

Hoe gaat opslag in DNA-vorm precies in zijn werk? Een DNA-streng is een polymeer, opgebouwd uit een opeenvolging van vier nucleotiden. Die worden aangeduid met de letters A, T, C en G (oftewel de organische moleculen Adenine, Thymine, Cytosine en Guanine). Het werkt in principe op dezelfde manier als de enen en nullen in een digitale bitstream. Kunnen we informatie digitaliseren, dan kunnen we het dus ook vertalen naar een nucleotiden-reeks. Vervolgens wordt die genetische code in een laboratorium gesynthetiseerd; de moleculen worden opgebouwd. Hierbij gaat het om kunstmatig gecreëerd materiaal, er komt geen levend organisme aan te pas. De strengen worden opgeslagen in een vloeibare oplossing en kunnen vervolgens relatief gemakkelijk worden bewaard en ver-

'Voorlopig is DNA-opslag alleen geschikt als een extreem duurzame back-up voor data die niet voortdurend geraadpleegd hoeft te worden'

'Onderzoekers van Microsoft verwachten dat het hooguit een decennium zal duren voordat de techniek echt klaar is om de wereld te veroveren'

voerd. Het maken van kopieën is snel en bovenal goedkoop: onder de juiste chemische omstandigheden kopieert DNA zichzelf, net als in het menselijk lichaam gebeurt.

Data opsplitsen

In de praktijk is het vertalen van een digitale code naar de juiste nucleotidenreeks echter iets ingewikkelder dan het op het eerste gezicht lijkt. Een DNA-streng mag bijvoorbeeld niet te lang zijn, en herhaling van te vaak dezelfde nucleotide achter elkaar vormt een risico op breuken. Om dit te voorkomen wordt één stuk data opgesplitst in tal van kleinere sliertjes DNA.

Ieder sliertje wordt voorzien van een zogenaamde *primer*. Die primer geldt als unieke identifier voor dat stuk data, en bevat ook aanvullende metadata over bijvoorbeeld de volgorde waarin de sliertjes gezet moeten worden om de data uit te lezen. Ieder sliertje komt honderden keren voor in het DNA-mengsel en bevat bovendien veel overlap met het voorgaande en opvolgende stukje bitstream. Op die manier draagt genetische data eigen back-ups mee en wordt de kans op lees- en kopieerfouten drastisch verminderd.

Hindernissen

De techniek om data in DNA op te slaan gaat met sprongen vooruit. Maar om het ook echt in de praktijk te kunnen gebruiken, moeten nog verschillende obstakels overwonnen worden. Het is hoofdzakelijk nog een heel langzaam proces. Het duurt nu vele uren om een bestand te openen (anders gezegd: het ontcijferen en digitaal weergeven van de nucleotidenreeks).

Dit weegt niet op tegen de microseconden die een flash drive voor dezelfde taak nodig heeft. De verwachting is dat er in de toekomst snelheden tot wel 100 Gb/sec behaald gaan worden. Tot die tijd is DNA-opslag alleen geschikt als een permanente, extreem duurzame back-up voor data die niet voortdurend geraadpleegd hoeft te worden. Denk bijvoorbeeld aan de kennis die niet verloren mag gaan bij een grote klimaat-ramp of wereldoorlog.

Maar de belangrijkste drempel is het prijskaartje. De kosten voor het deco-

deren en synthetiseren van DNA mogen dan al rap dalen, het is nog steeds een prijzige aangelegenheid. De bedragen voor het synthetiseren van enkele MB's aan data lopen al gauw op tot een paar duizend dollar. Daar durft een archief in de culturele sector nog niet eens van te dromen.

Zelfs wanneer het schrijven van DNA betaalbaar wordt, zullen de serverkasten nog niet direct aan de straat gezet worden. Transcoderen van digitale naar genetische data is één ding, maar het vervangen van de huidige infrastructuur in de datacenters is eveneens een immense klus. Commerciële leveranciers zullen er niet zomaar aan beginnen zodra de techniek ook buiten de wetenschappelijke wereld voorhanden komt.

Kwestie van tijd

Volgens schattingen gepubliceerd door *Nature* zal het wereldwijde digitale archief in 2040 zo extreem veel data bevatten, dat de opslag ervan op huidige flash memorysticks honderd keer meer silicium van microchip-kwaliteit nodig heeft dan er naar verwachting op aarde voorradig is. Dat DNA data-opslag voor de oplossing gaat zorgen lijkt gelukkig slechts een kwestie van tijd. De onderzoekers van Microsoft verwachten dat het hooguit een decennium zal duren voordat de techniek echt klaar is om de wereld te veroveren. Wordt de informatiefprofessional straks ook bio-chemicus?

Nog meer toepassingen

DNA is breder inzetbaar dan als opslagmedium alleen. Neem een computer. Op basis van DNA is die in potentie sneller dan de huidige apparatuur. En bovendien een stuk energiezuiniger: uitwisselen van data is hierbij chemisch van aard en er wordt geen elektriciteit verbruikt.

Verder onderzoekt Harvard Medical School of levende cellen via DNA geprogrammeerd kunnen worden om zelfstandig informatie te registreren. Een soort levende micro-camera dus die biologische veranderingen kan vastleggen.